

# Trends in the Click Probability

STATCOM

## Executive Summary

### Questions of Interest

- What are the factors that increase the probability of a subscriber clicking on any link in a newsletter?
  - Does the time of day the newsletter is sent affect the click rate?
  - Does the email subject affect the click rate?
  - Does the content of the newsletter (word count, number of links, number of pictures) affect the click rate?
- How has the COVID pandemic affected the click rate?

### Statistical Analysis

- Barplots to visually display trends
- Statistical model to examine how factors interact
- Results from the model to confirm the trends shown in the barplots

### Takeaways

- Because the click probability is already quite low, changing the time of day or subject length would not affect the click probability as much as it would affect the open probability. Thus, we recommend trying to increase the open rate before trying to change the click rate with those factors.
- Before the pandemic, there were less clicks if the newsletter was sent around 9 am and more clicks if the newsletter was sent around 1 pm. However, during the pandemic, the time of day the newsletter was sent didn't affect the click rate.
- A subject length of 70 characters leads to the most clicks.
- A higher word count is associated with less clicks.
- More text or image links lead to more clicks.
- The optimal number of clickable pictures appears to be 8. More clickable pictures is not necessarily better.

# Trends in the Click Probability

## Overview of the Data Used

There are 16,291 unique subscribers in the data set, with 622,614 observations.

We focused on the probability that a given subscriber will click on at least one link in a given newsletter. In addition to the variables used in the open probability model,

- Trend over the date newsletter was sent out (2019-01-01 to 2020-12-31).
- Whether newsletter was sent before or after start of the COVID pandemic on 2020-03-12.
- Time of day newsletter was sent out (6:30 am to 8:40 pm).
- Length of subject by number of characters.

we also look at the following factors related to the content within the newsletter:

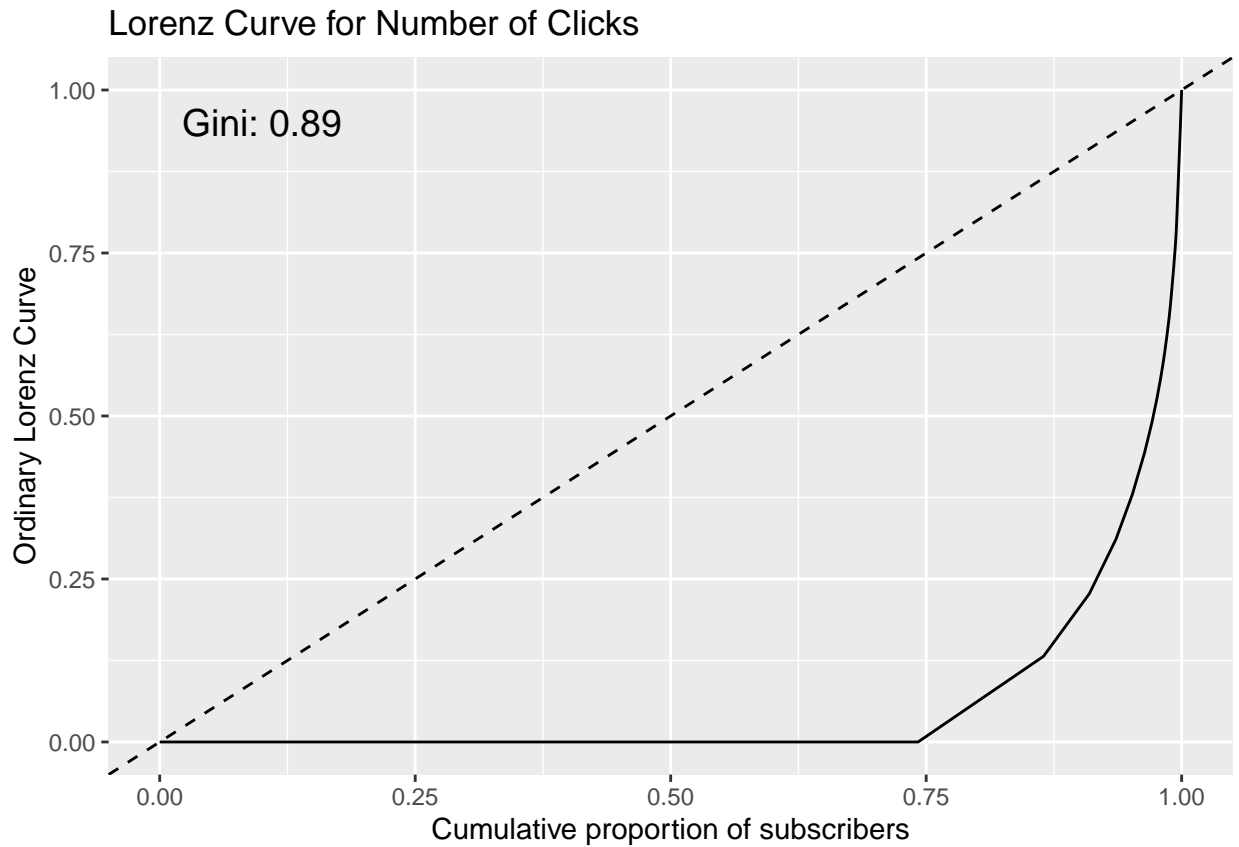
- Number of words in newsletter.
- Number of links (text or image) in newsletter.
- Number of clickable pictures in newsletter.
- Number of unclickable pictures in newsletter.

Below are 10 sample observations from the dataset, containing the four “content” variables mentioned above. Each row corresponds to one of the 622,614 newsletter-subscriber pairs. The last variable, clicks, is the response variable of interest. 1 indicates that the subscriber clicked on any link in a given newsletter; 0 indicates that the subscriber didn’t click on any link in a given newsletter.

num_words	num_links	num_clickable_pics	num_unclickable_pics	clicks
347	23	6	0	0
410	23	6	1	0
402	28	8	0	0
512	30	9	0	0
333	20	6	0	0
357	30	13	0	0
382	26	6	0	0
484	31	7	3	0
484	31	7	3	0
513	27	8	0	0

## Lorenz Curve

The Gini Index ranges from 0 to 1, with 1 being perfect inequality. In this case, the distribution of clicks among the subscribers seems unequal; according to the curve, only 25% of subscribers click at all, and the top 10% of people account for 75% of clicks.



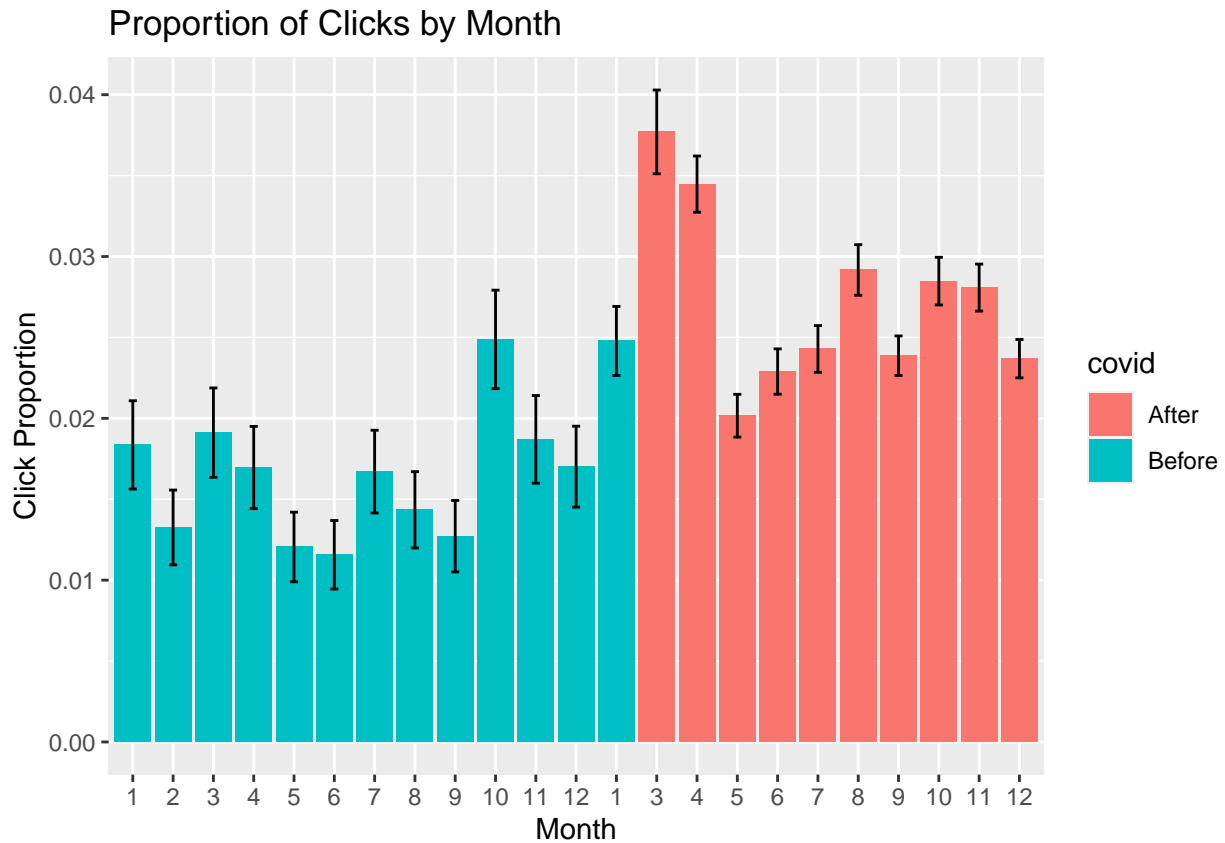
## Overview of Analysis

The generalized additive mixed model uses a random sub-sample of 1,629 subscribers (63,897 observations) so it can finish in a reasonable amount of time.

All plots show 95% confidence intervals of the estimates; two standard errors above and below the estimates are indicated on top of the bars in the barplots and by the dashed lines or shading in the line graphs. The true value can be expected to lie within two standard errors from the estimate.

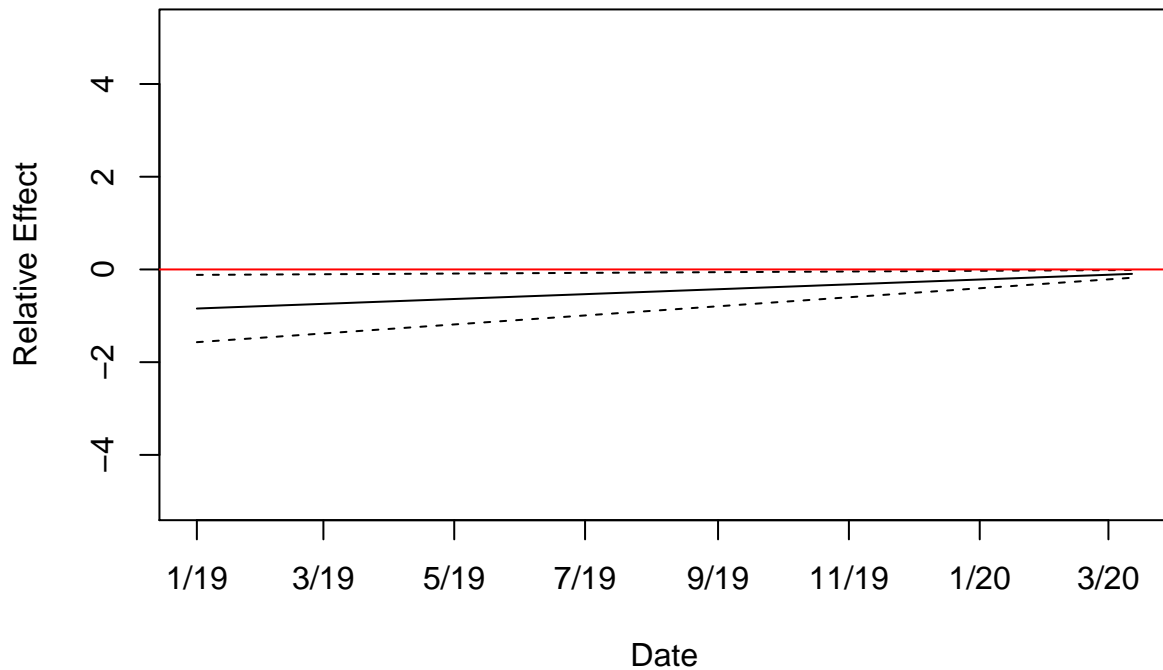
## Trend over Date

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given the month the newsletter was sent to them.



The following plot shows the relative effect of the date the newsletter is sent out on the click probability (a negative relative effect corresponds to a decrease in probability, and a positive relative effect corresponds to an increase in probability) before the pandemic. There is a slight increase over time leading up to the pandemic.

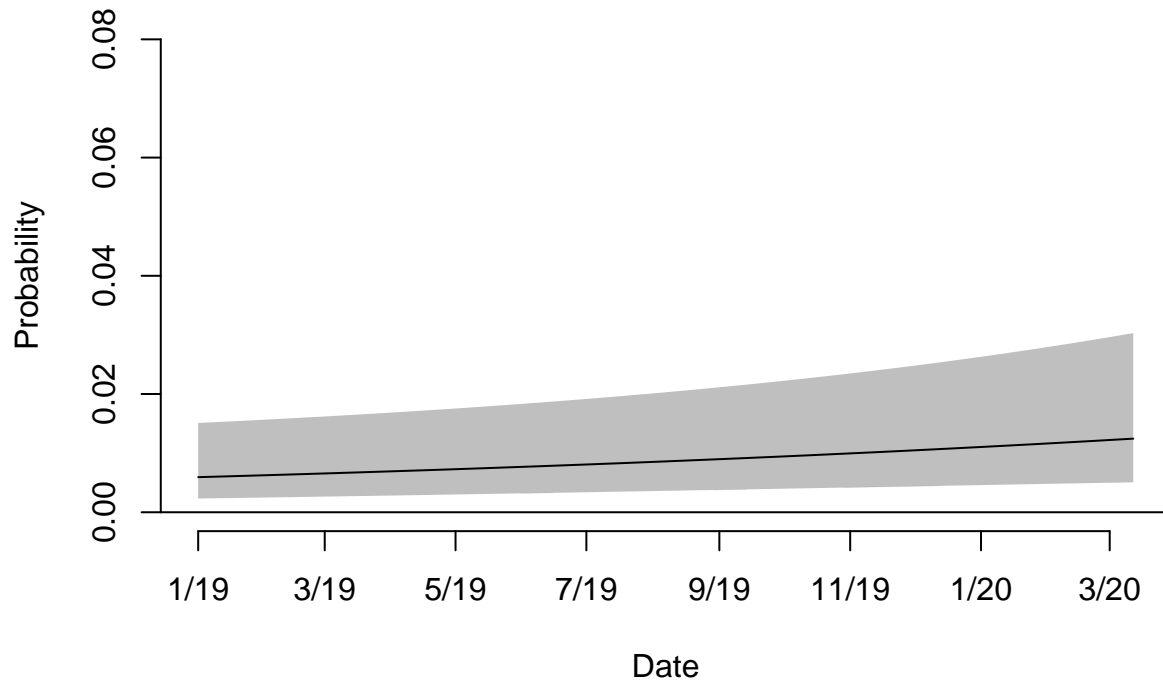
## Click Probability over Time Before COVID



The above plot shows the partial effect of the date alone, without considering other covariates. The following plot shows the actual estimated probabilities over time under the following specific scenario:

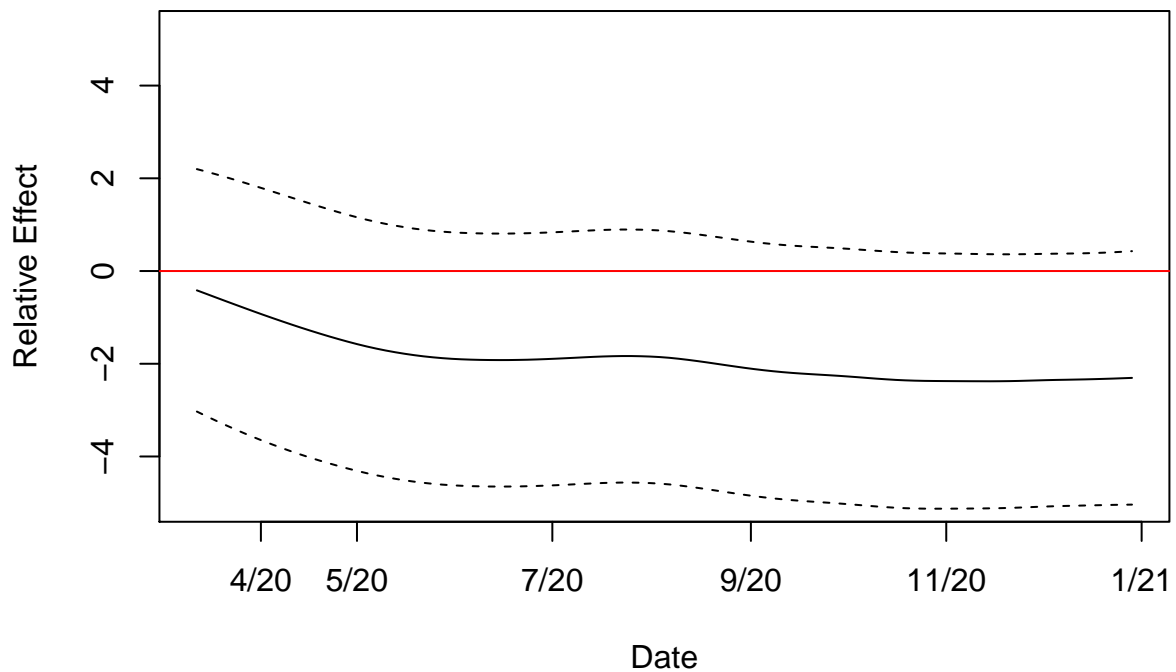
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

### Click Probability over Time Before COVID



After the pandemic, there is a downward trend in click probability.

### Click Probability over Time During COVID

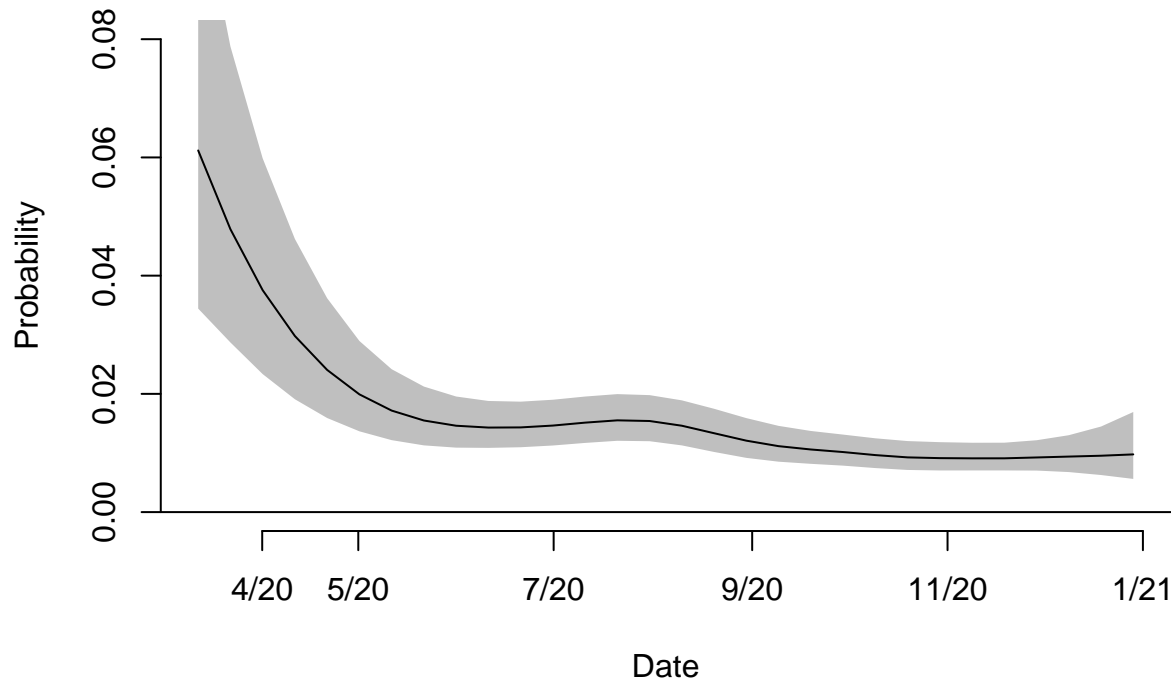


The above plot shows the partial effect of the date alone, without considering other covariates. The following plot shows the actual estimated probabilities over time under the following specific scenario:

- The newsletter was sent out at 10:30 am.

- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

### Click Probability over Time During COVID



Below shows the relative effect of COVID (solid line), i.e. whether the newsletter was sent after the pandemic started. It appears that the click probability rises after the pandemic starts, but the effect is not significant (see the dashed standard error bars). However, COVID significantly affects how the click probability varies by date or hour of day the newsletter was sent.

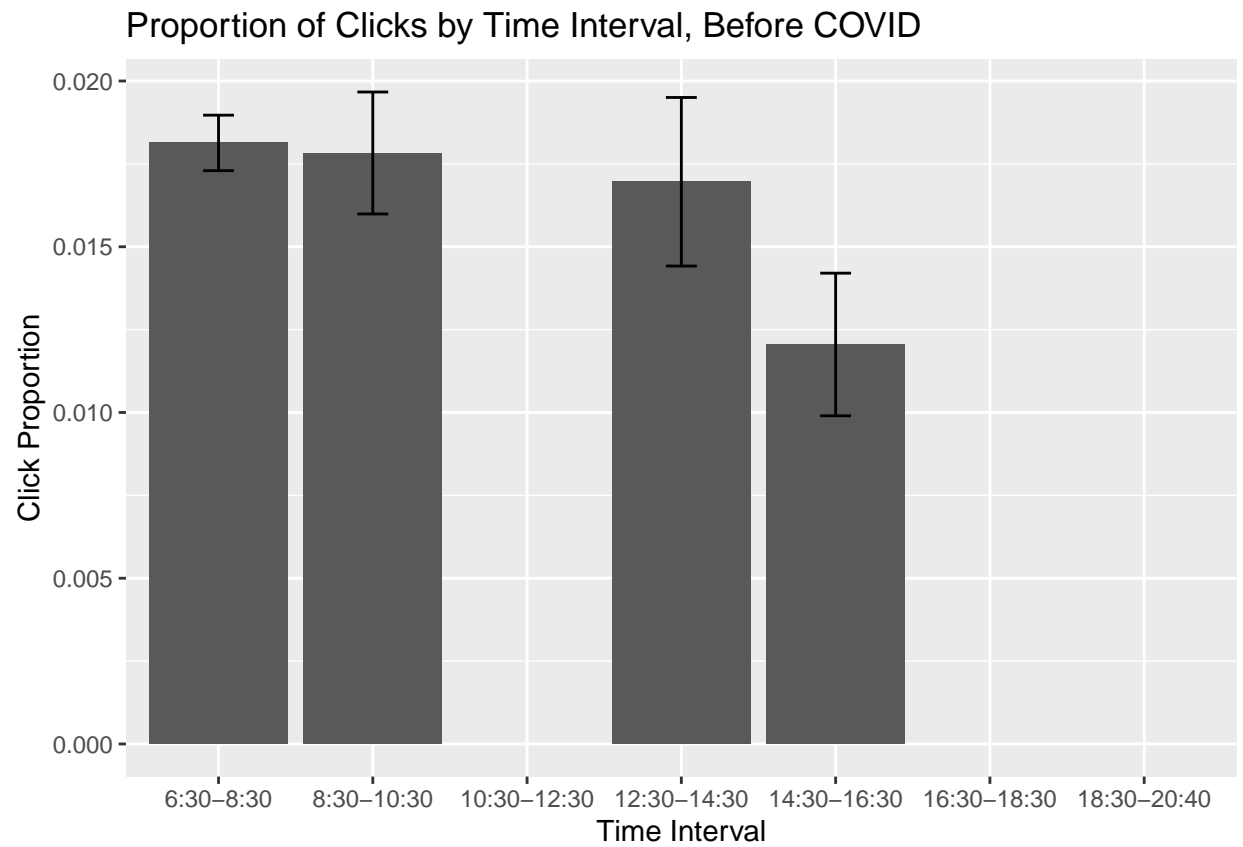
## Relative Effect of COVID



## Time of Day Trend, Before COVID

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter sent before the pandemic, given that the newsletter was sent to them within a specific time interval. There were no newsletters sent in three of the time intervals, so the bars are absent.

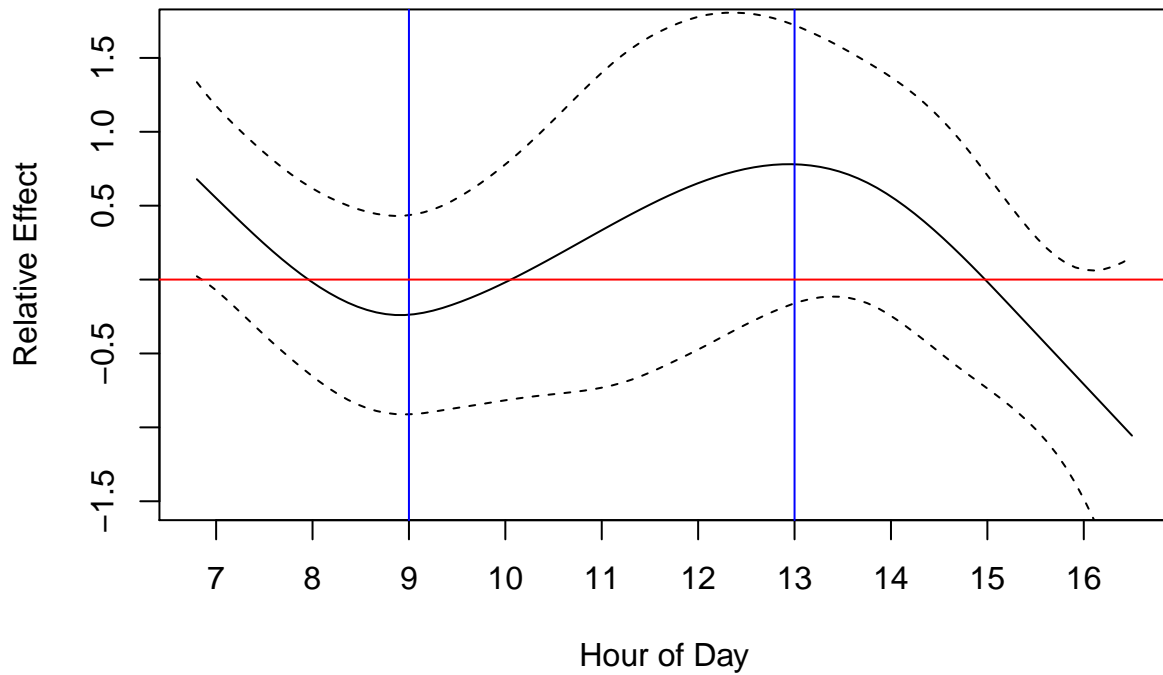




The following plot shows the relative effect of the time of day the newsletter is sent out on the click probability (a negative relative effect corresponds to a decrease in probability, and a positive relative effect corresponds to an increase in probability) before the pandemic.

It appears that there is a dip in the click probability at about 9 am, whereas there is a rise at about 1 pm.

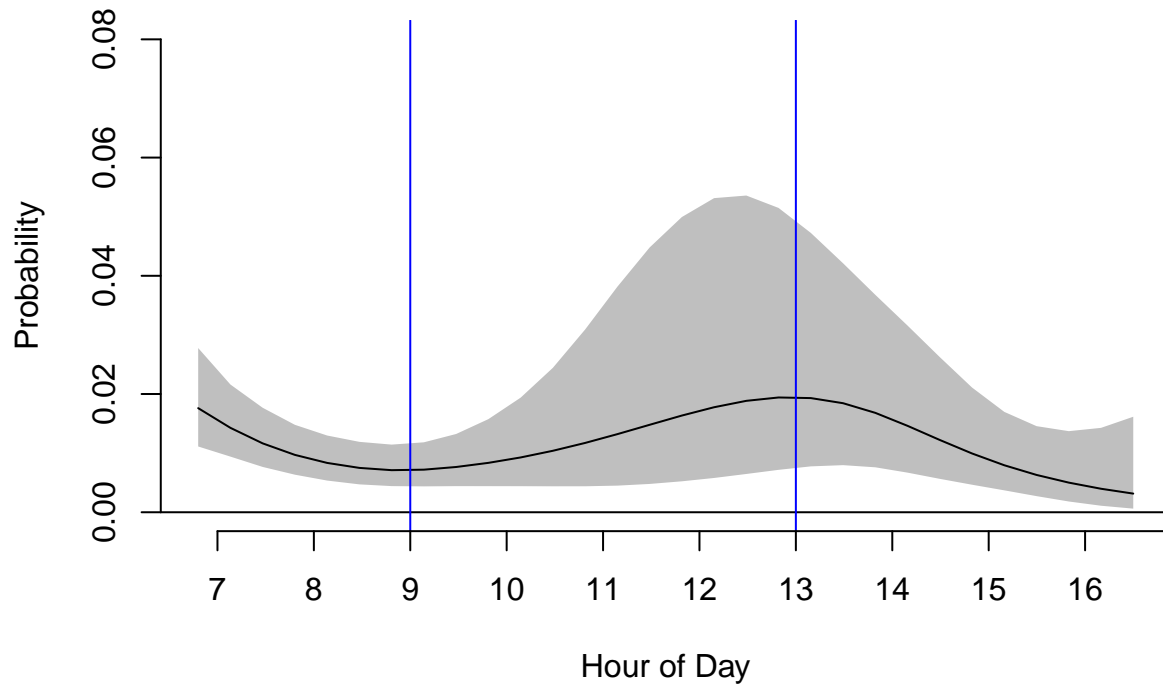
## Effect of Hour of Day on Click Probability, Before COVID



The above plot shows the partial effect of the time of day alone, without considering other covariates. The following plot shows the actual estimated probabilities by time of day under the following specific scenario:

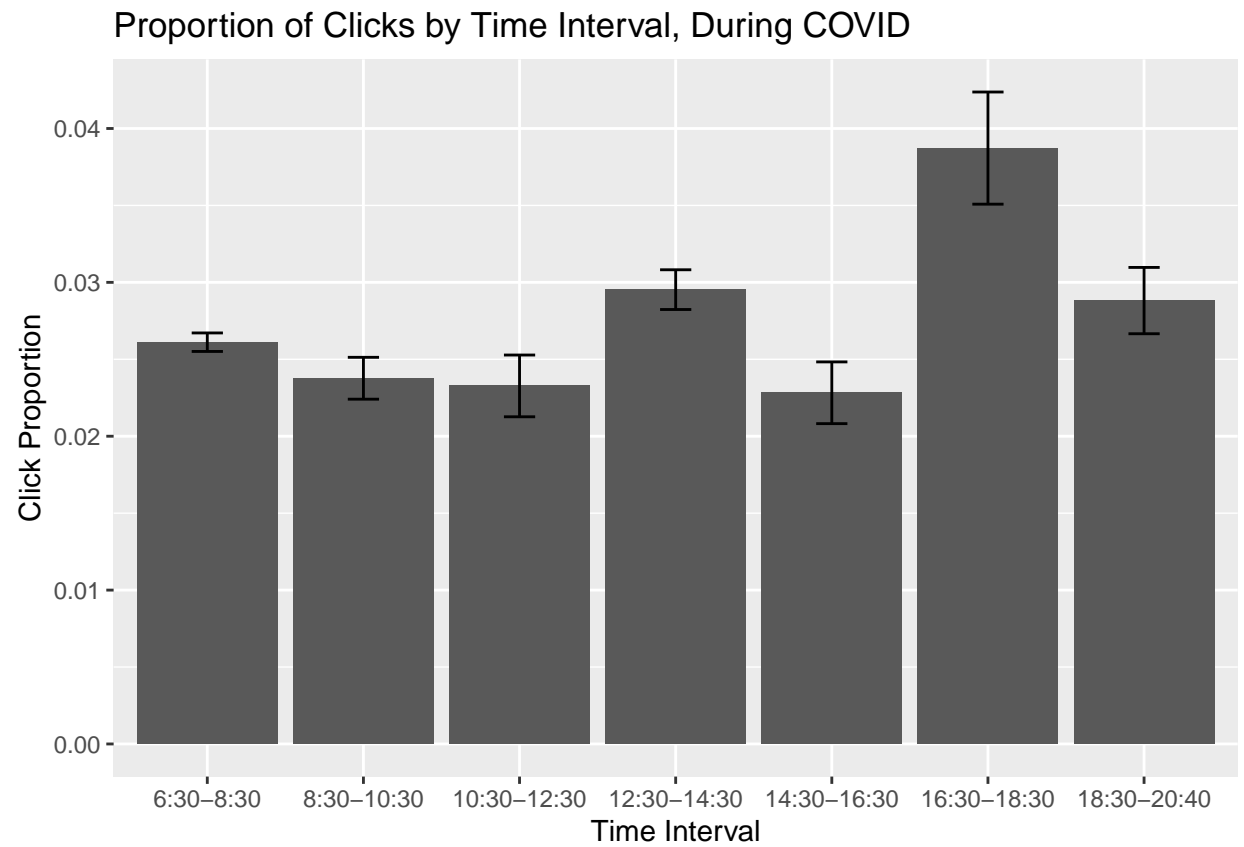
- The newsletter was sent out on December 1, 2019.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

## Click Probability vs. Hour of Day, given other covariates



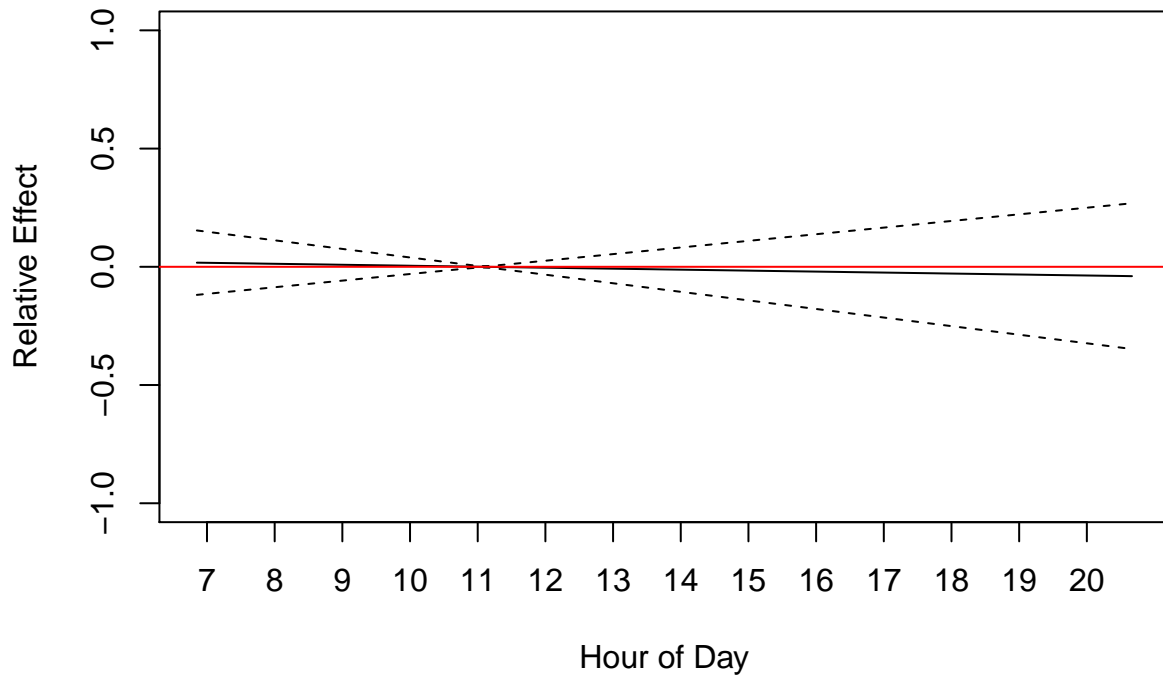
## Time of Day Trend, During COVID

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter sent during the pandemic, given that the newsletter was sent to them within a specific time interval.



The following plot shows the relative effect of the time of day the newsletter is sent out on the click probability during the pandemic. During the pandemic, the time of day the newsletter is sent has no discernible effect on the click probability.

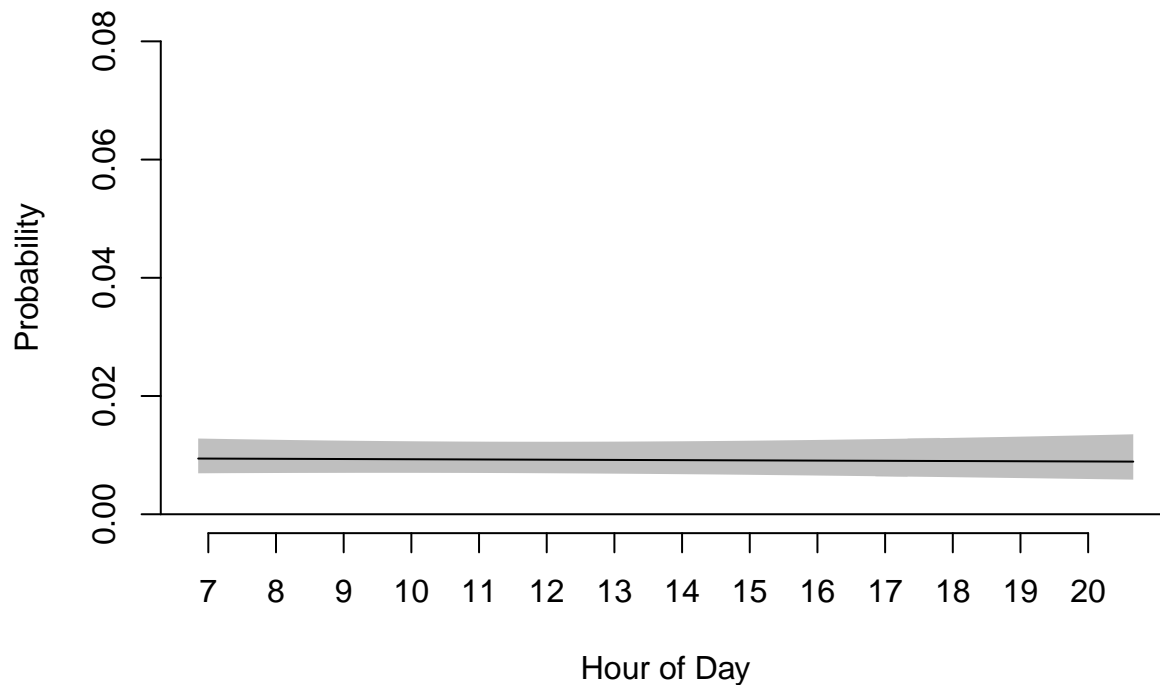
## Effect of Hour of Day on Click Probability, During COVID



The above plot shows the partial effect of the time of day alone, without considering other covariates. The following plot shows the actual estimated probabilities by time of day under the following specific scenario:

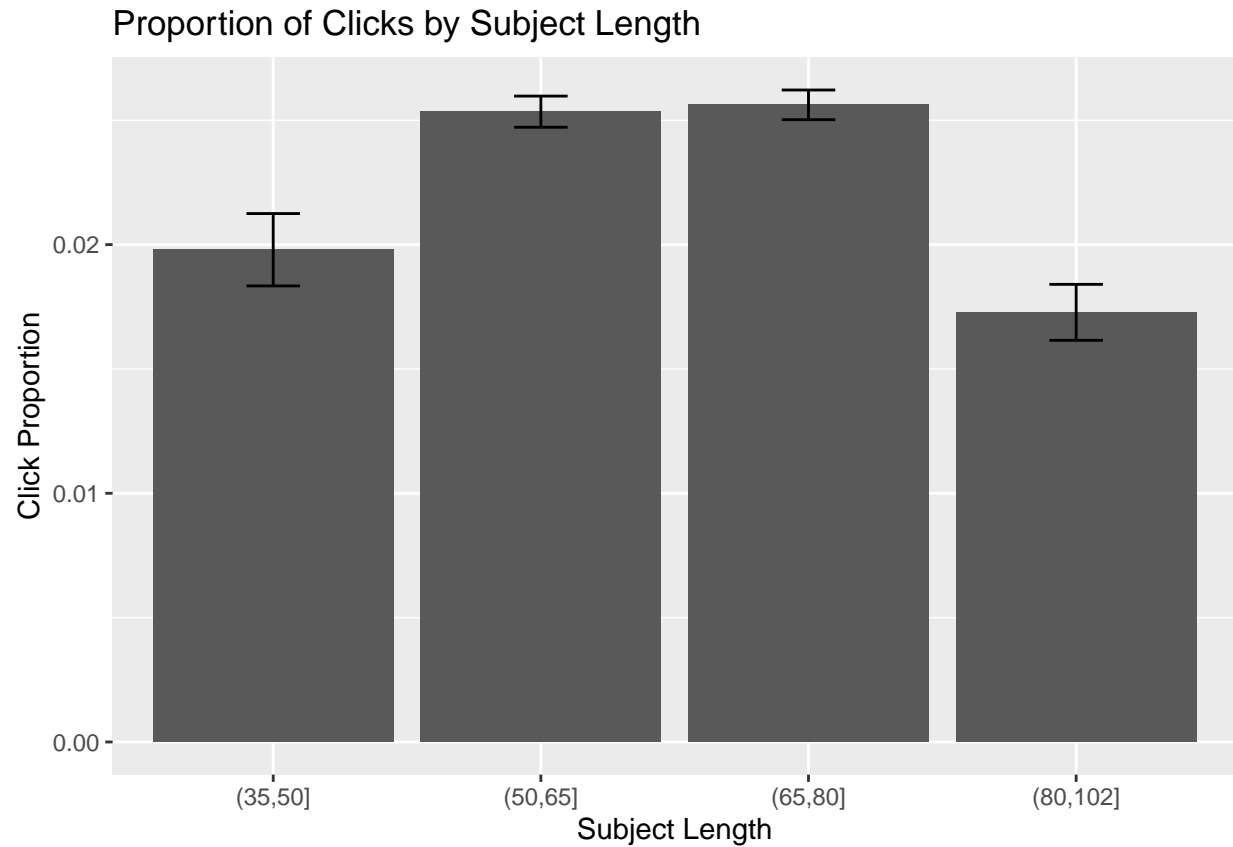
- The newsletter was sent out on December 1, 2020.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

### Click Probability vs. Hour of Day, given other covariates

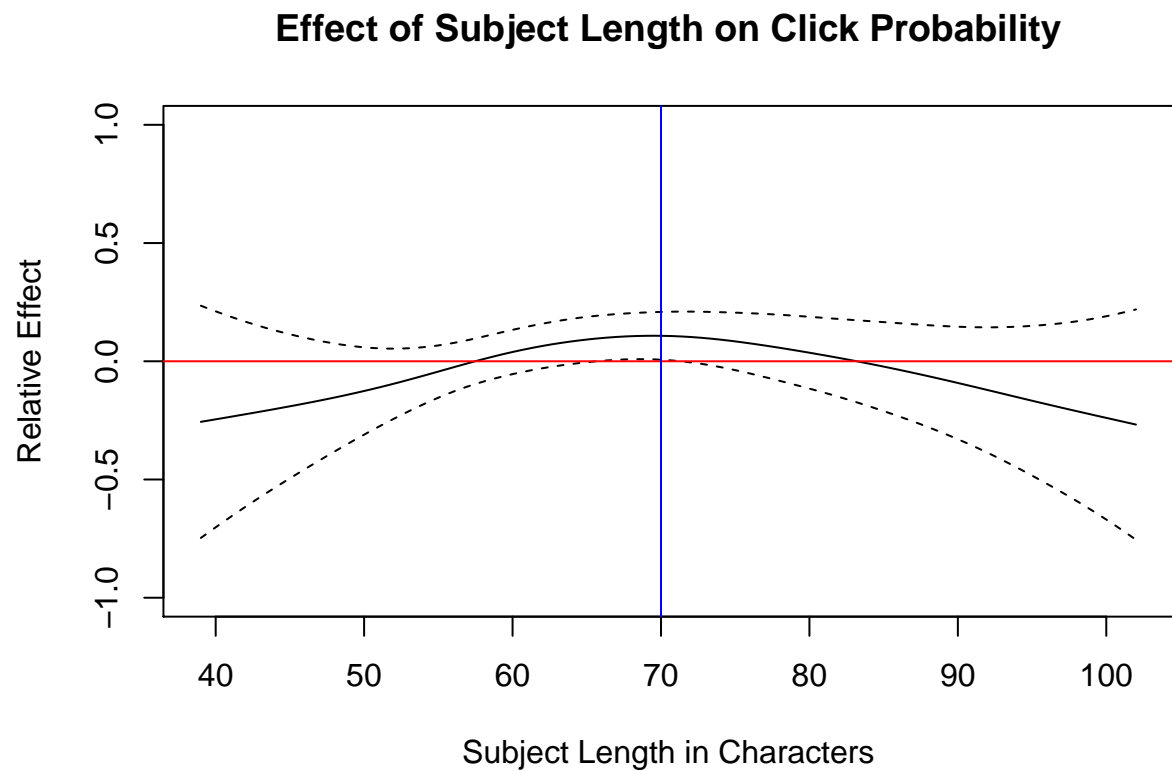


### Subject Length Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given that the subject length was within a specific interval.



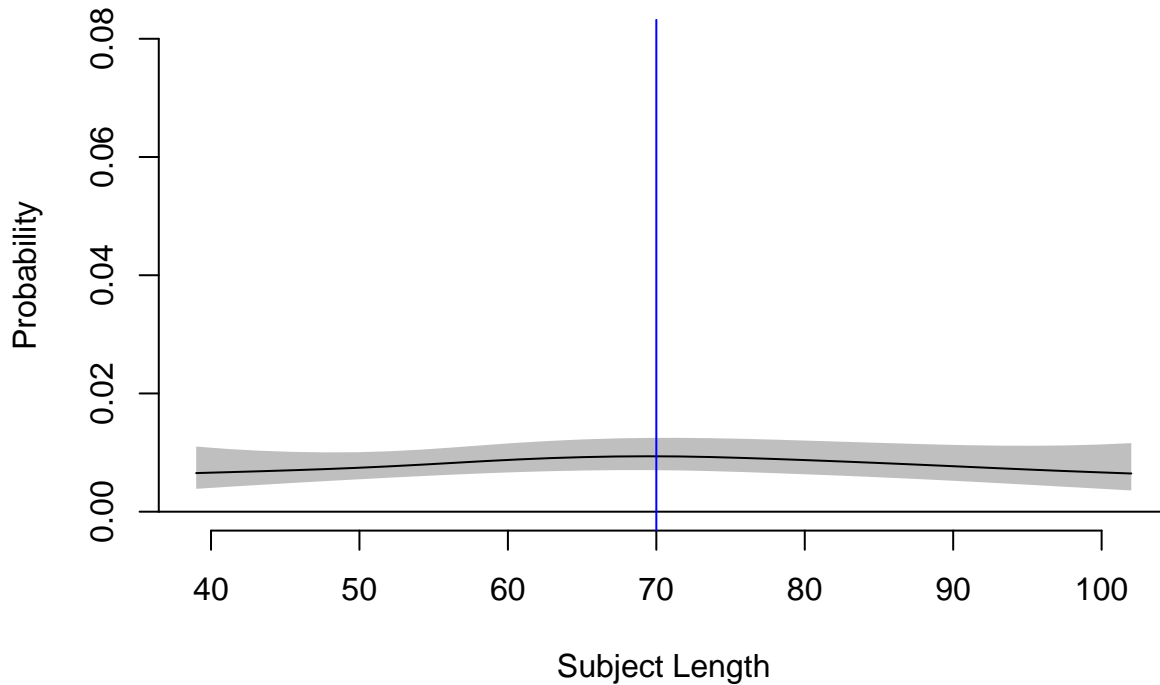
The following plot shows the relative effect of the subject length on the click probability. The maximum click probability corresponds to a subject length of approximately 70 characters.



The above plot shows the partial effect of the subject length alone, without considering other covariates. The following plot shows the actual estimated probabilities by subject length under the following specific scenario:

- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

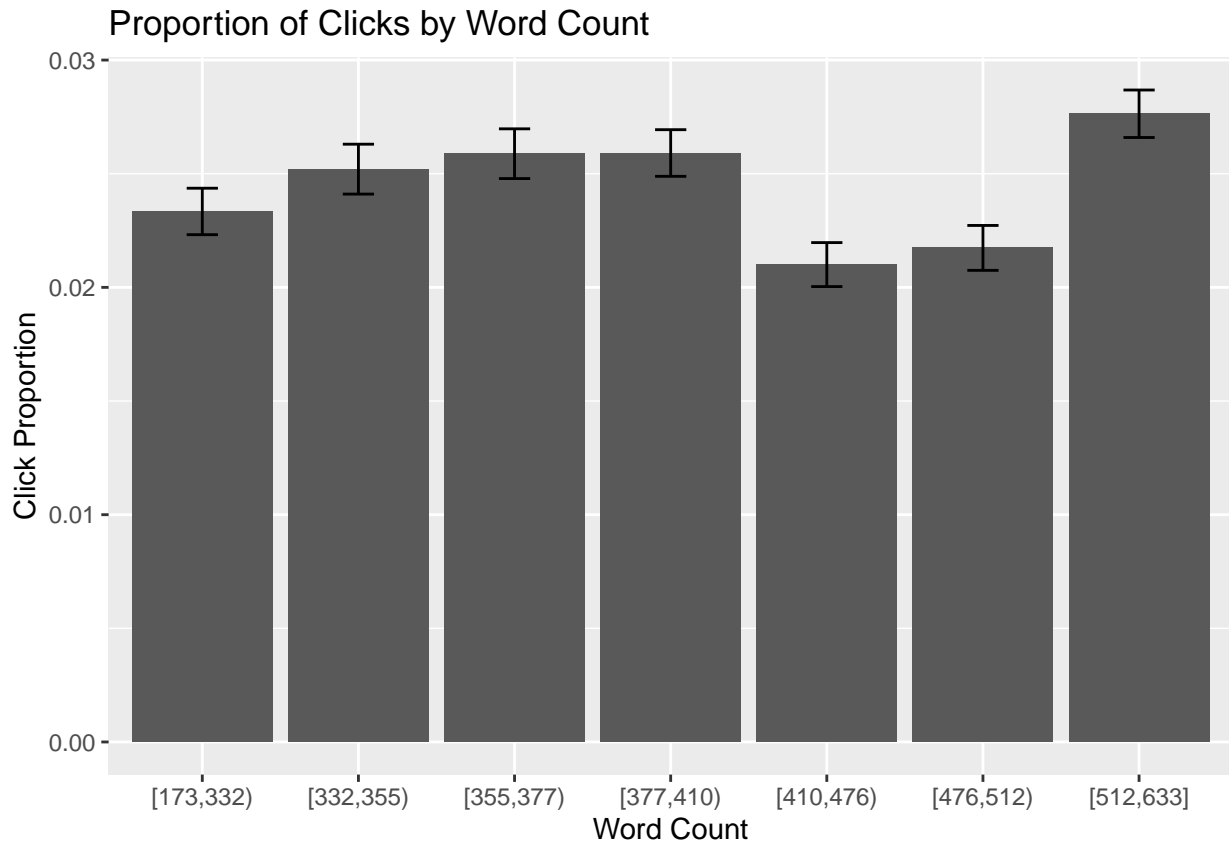
### Click Probability vs. Subject Length, given other covariates



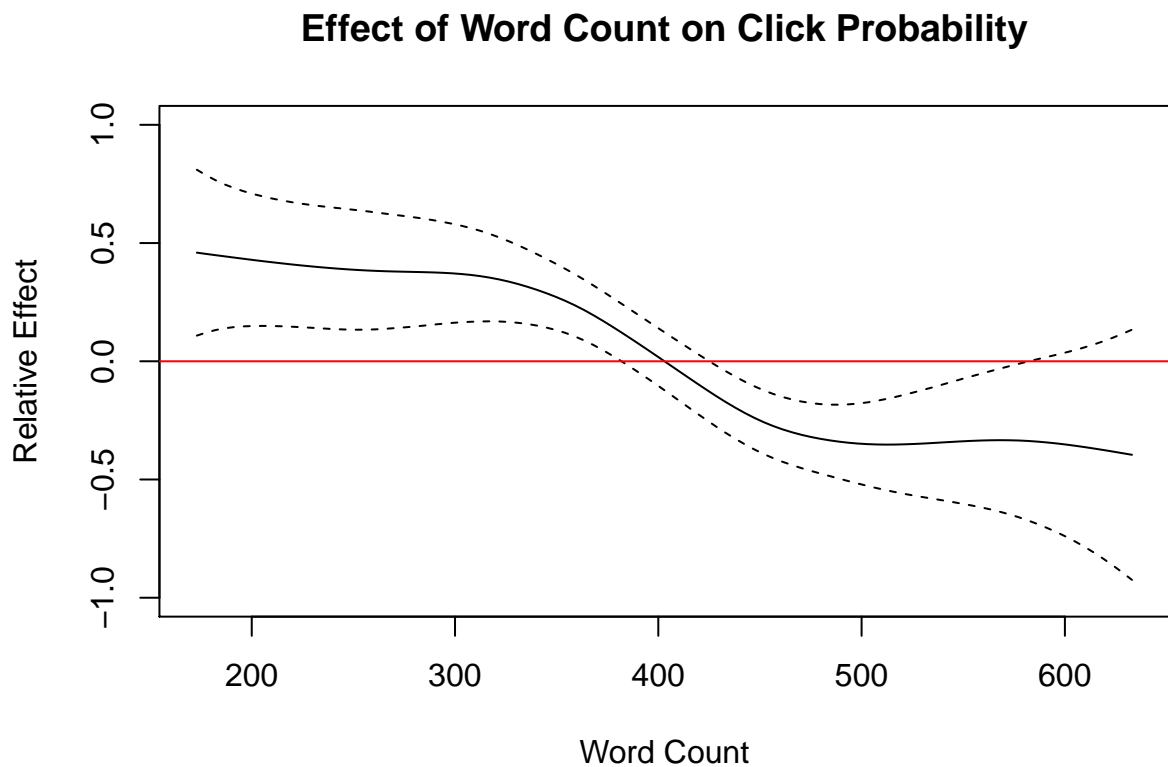
### Word Count Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given that the newsletter word count was within a specific interval.





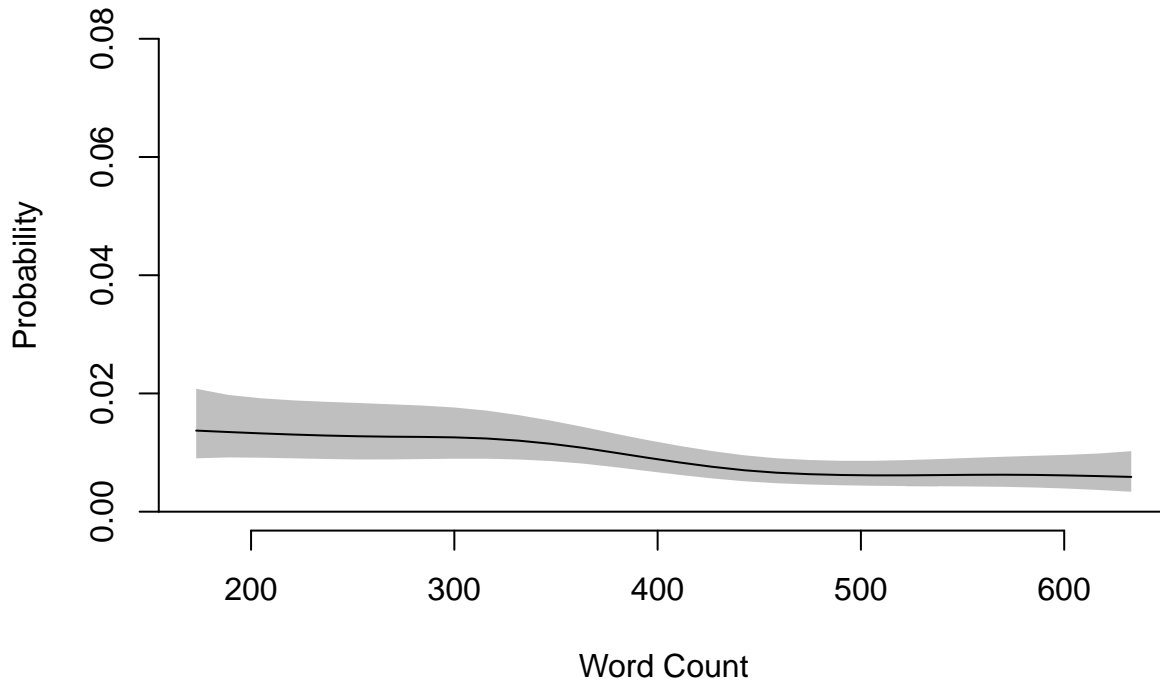
The following plot shows the relative effect of the word count on the click probability. There appears to be a downward trend in click probability as the word count increases.



The above plot shows the partial effect of the word count alone, without considering other covariates. The following plot shows the actual estimated probabilities by word count under the following specific scenario:

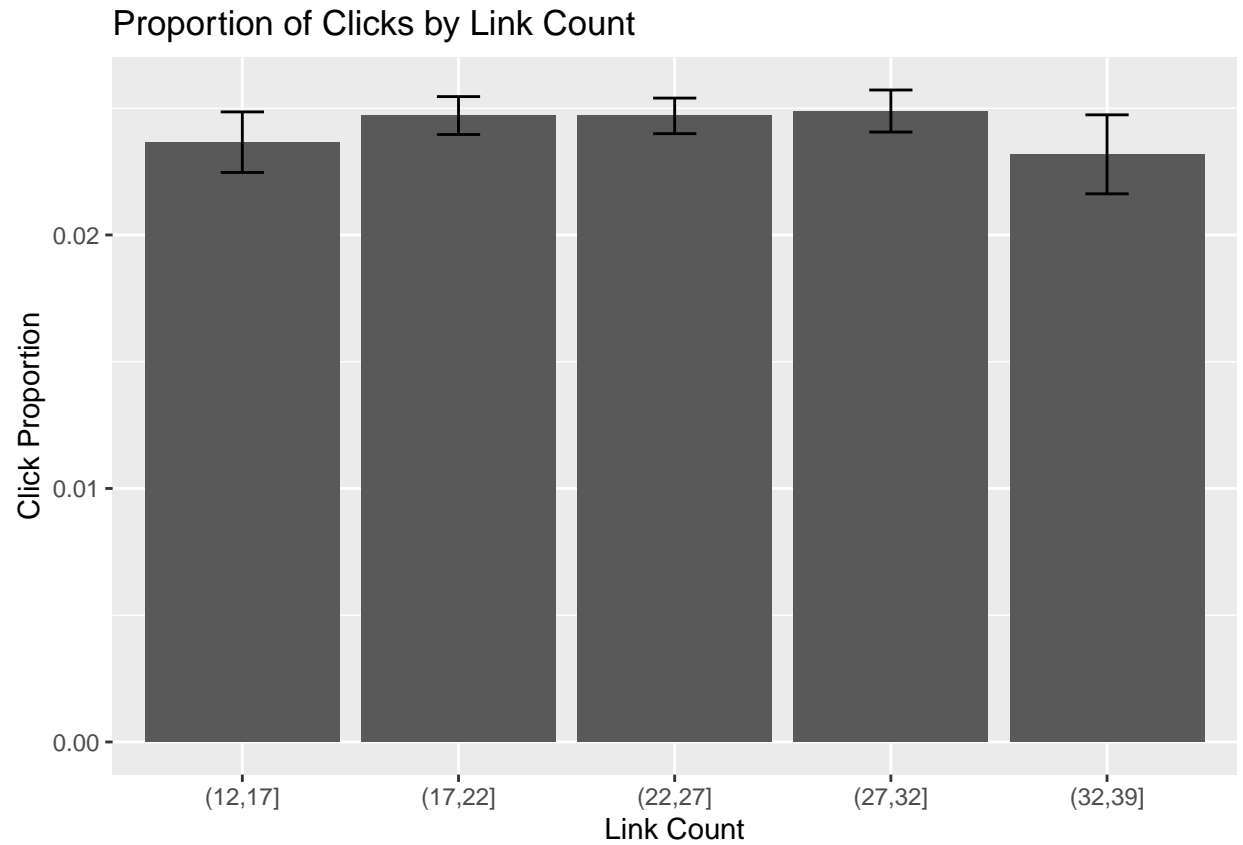
- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

### Click Probability vs. Word Count, given other covariates

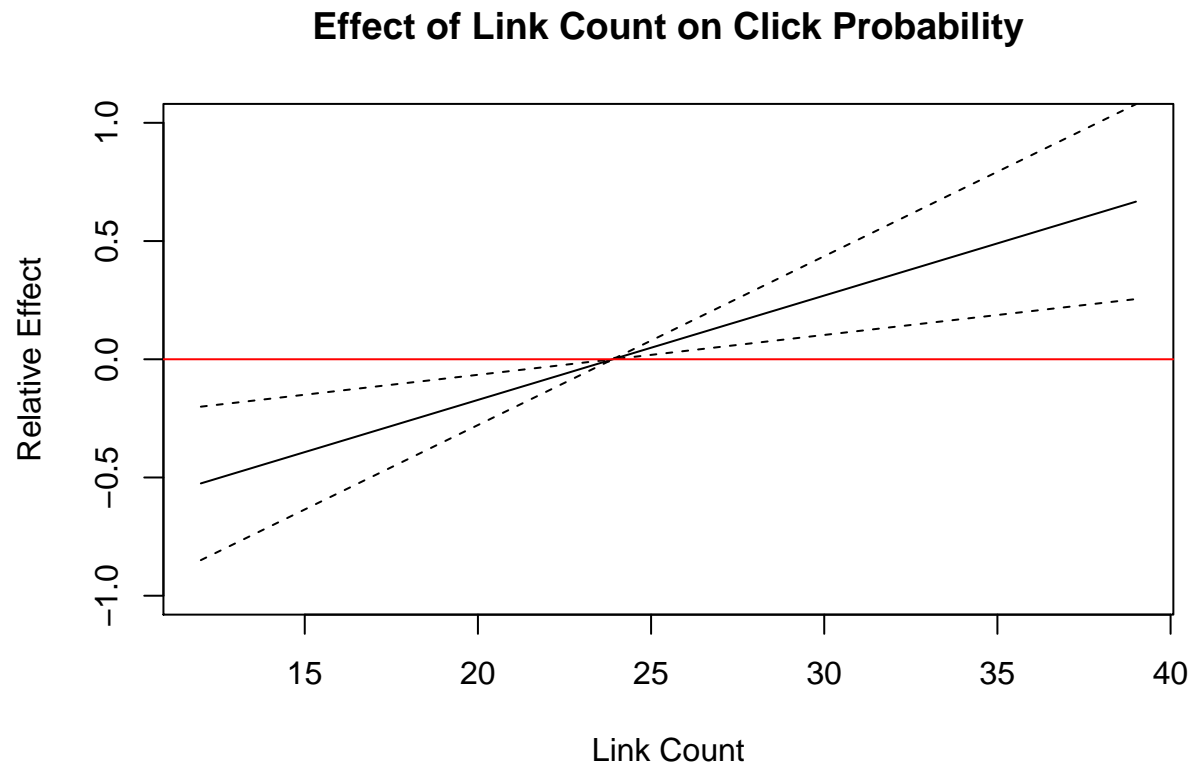


### Number of Links Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given that the newsletter link count was within a specific interval.



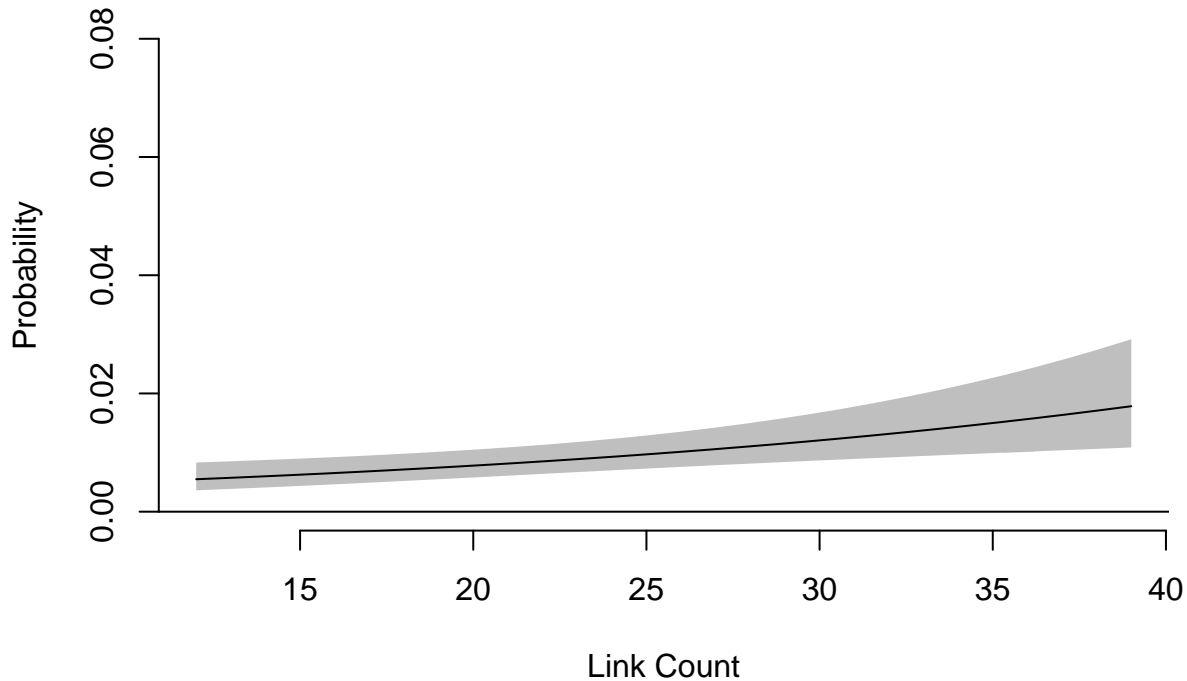
The following plot shows the relative effect of the link count on the click probability. More links is associated with a higher click probability.



The above plot shows the partial effect of the link count alone, without considering other covariates. The following plot shows the actual estimated probabilities by link count under the following specific scenario:

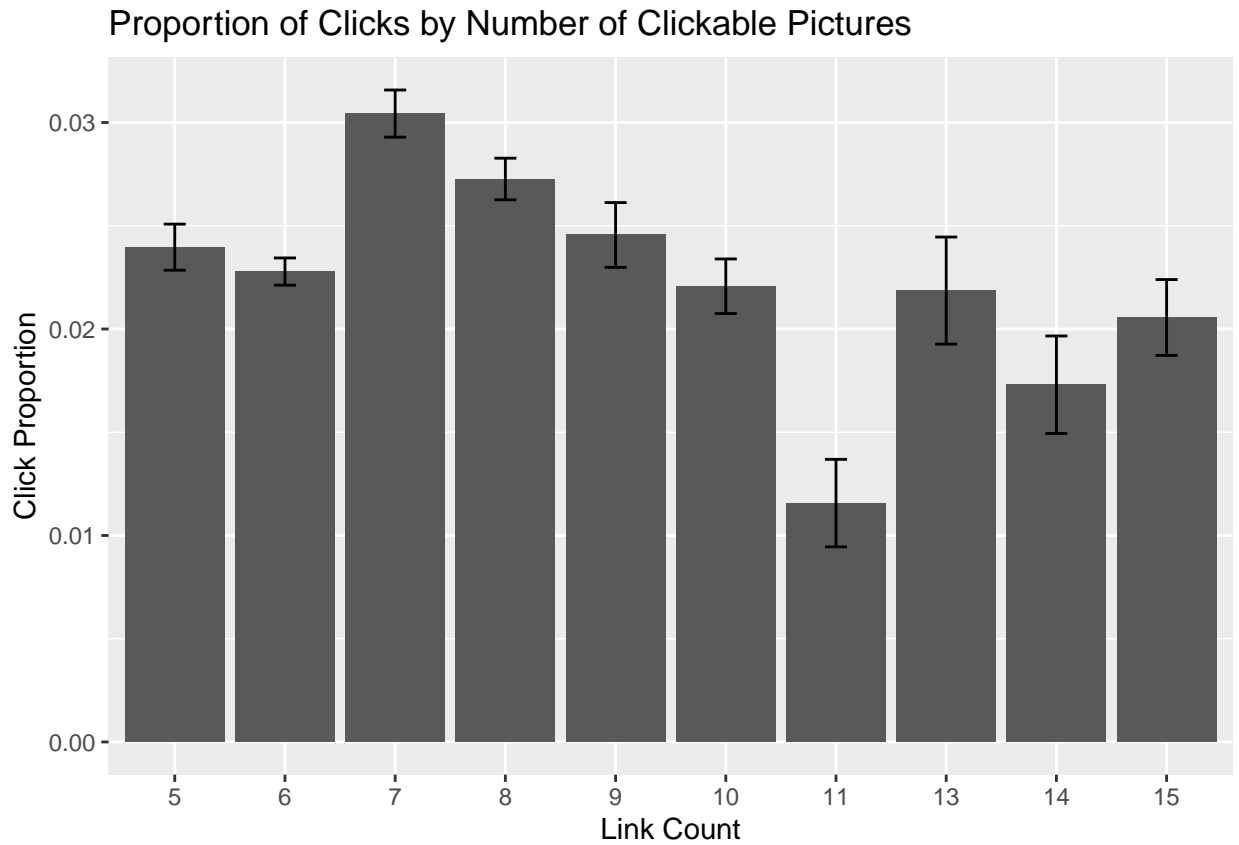
- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

### Click Probability vs. Link Count, given other covariates

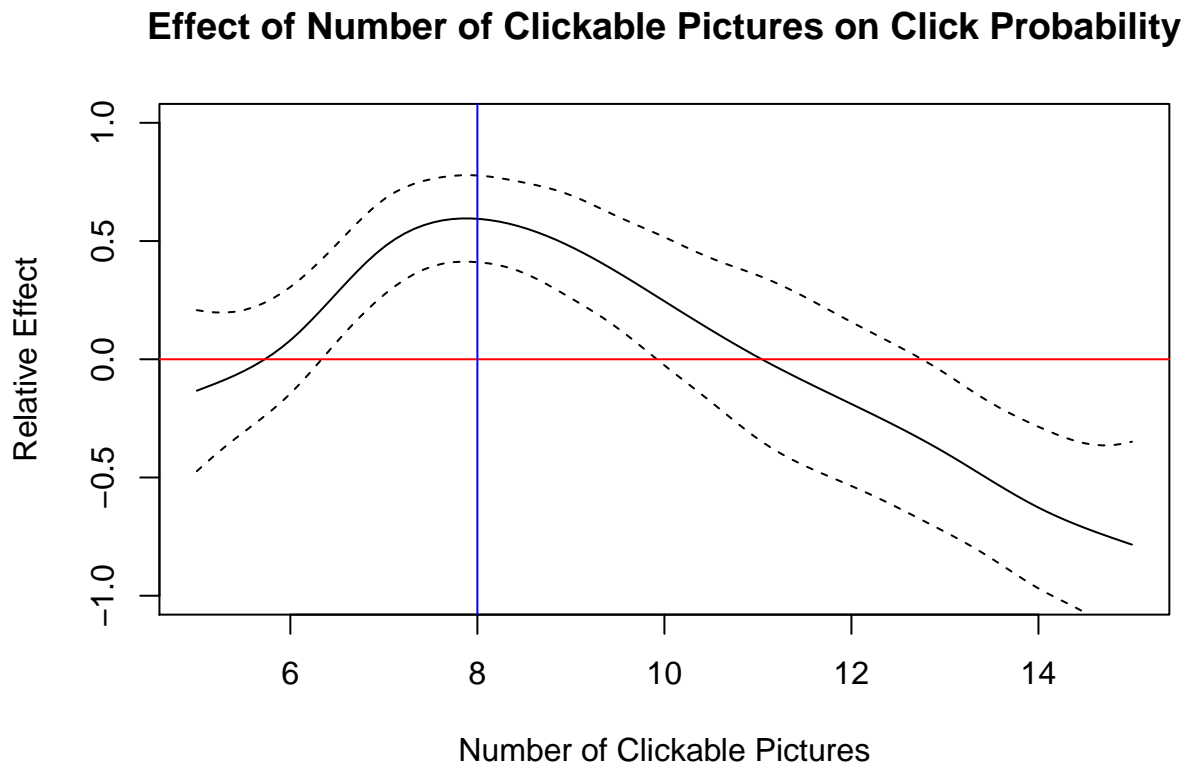


### Number of Clickable Pictures Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given the number of clickable pictures in the newsletter.



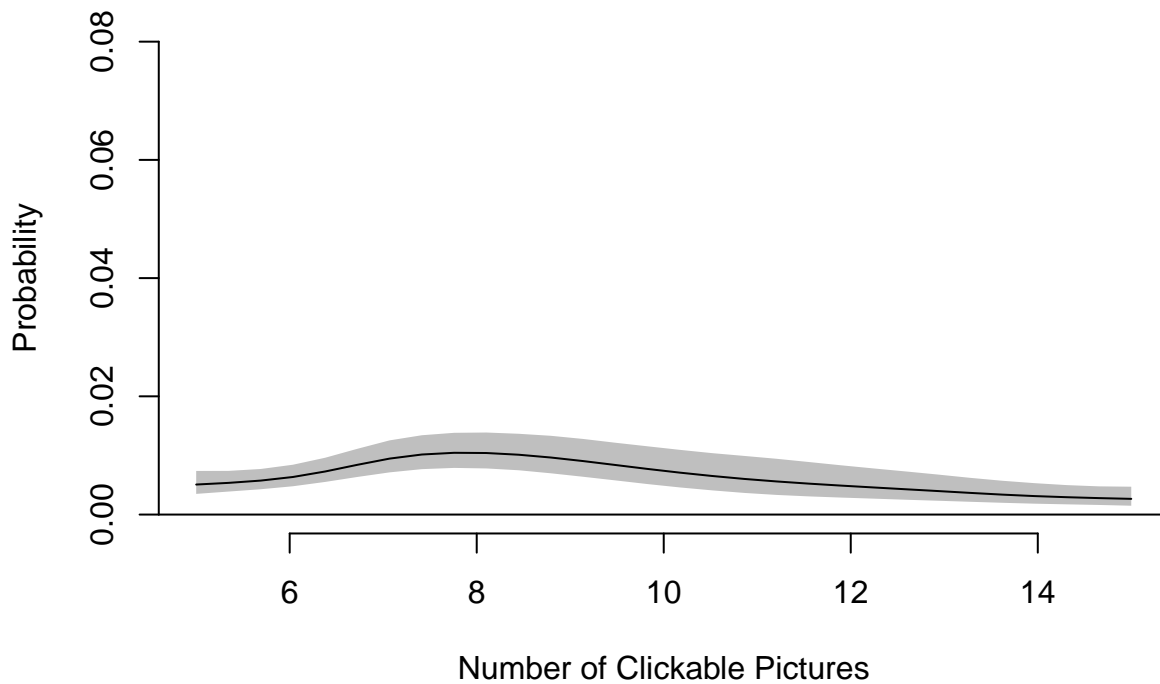
The following plot shows the relative effect of the number of clickable pictures on the click probability. The optimal number of clickable pictures appears to be 8.



The above plot shows the partial effect of the number of clickable pictures alone, without considering other covariates. The following plot shows the actual estimated probabilities by number of clickable pictures under the following specific scenario:

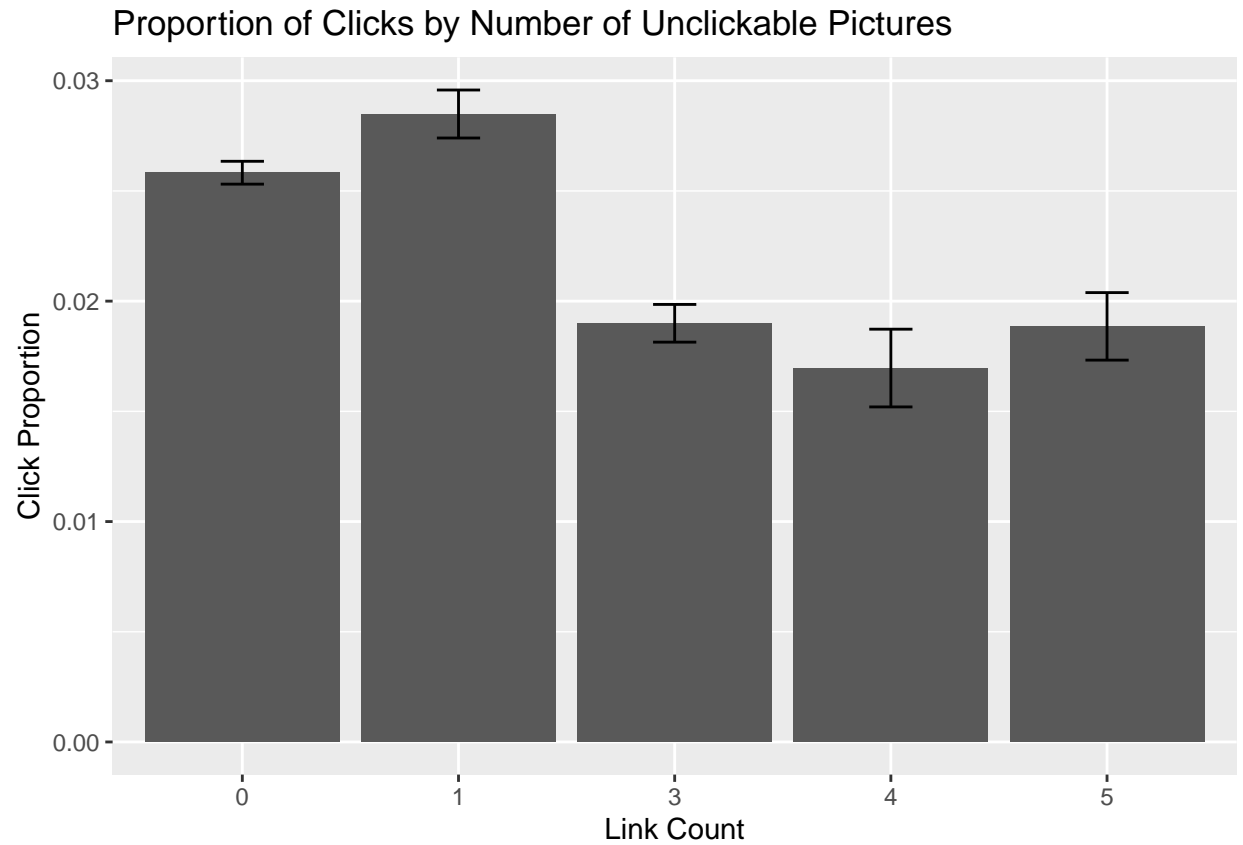
- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of unclickable pictures at 0.

### Click Probability vs. Number of Clickable Pictures

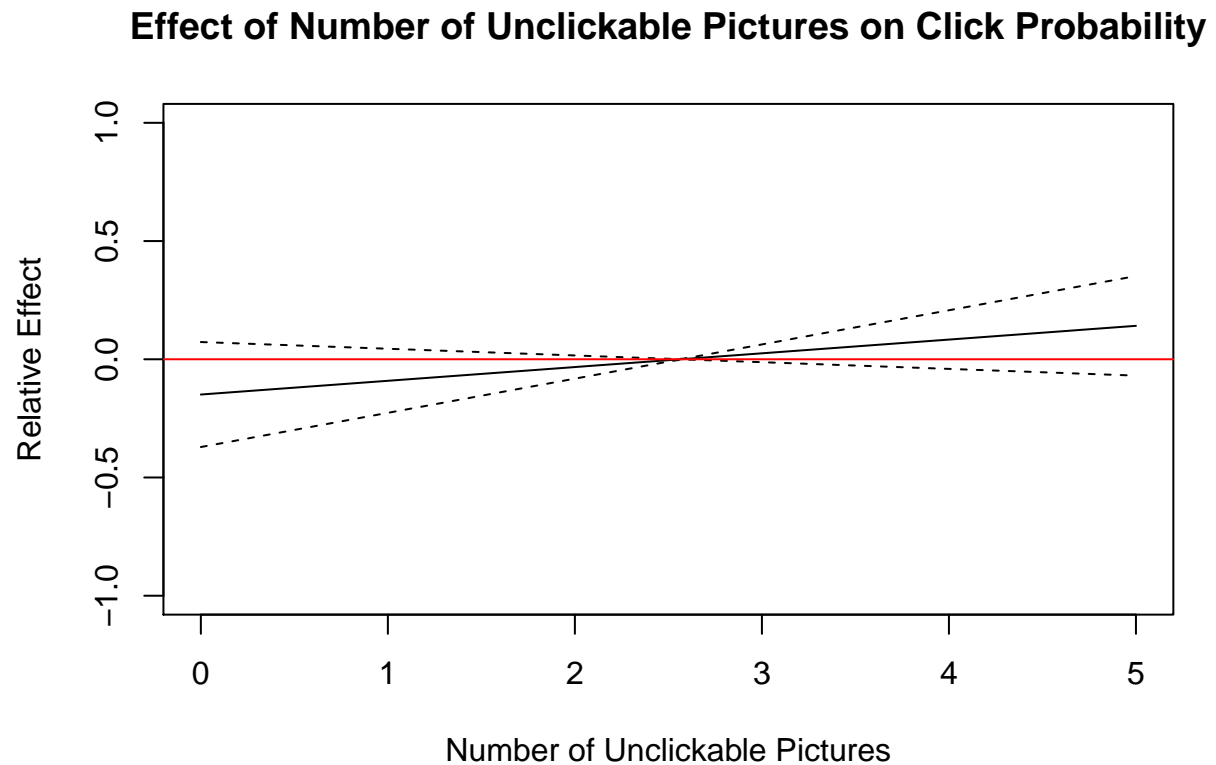


### Number of Unclickable Pictures Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given the number of unclickable pictures in the newsletter.



The following plot shows the relative effect of the number of unclickable pictures on the click probability. The number of unclickable pictures does not have a significant effect on click probability



The above plot shows the partial effect of the number of unclickable pictures alone, without considering other covariates. The following plot shows the actual estimated probabilities by number of unclickable pictures under the following specific scenario:

- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.

### Click Probability vs. Number of Unclickable Pictures

